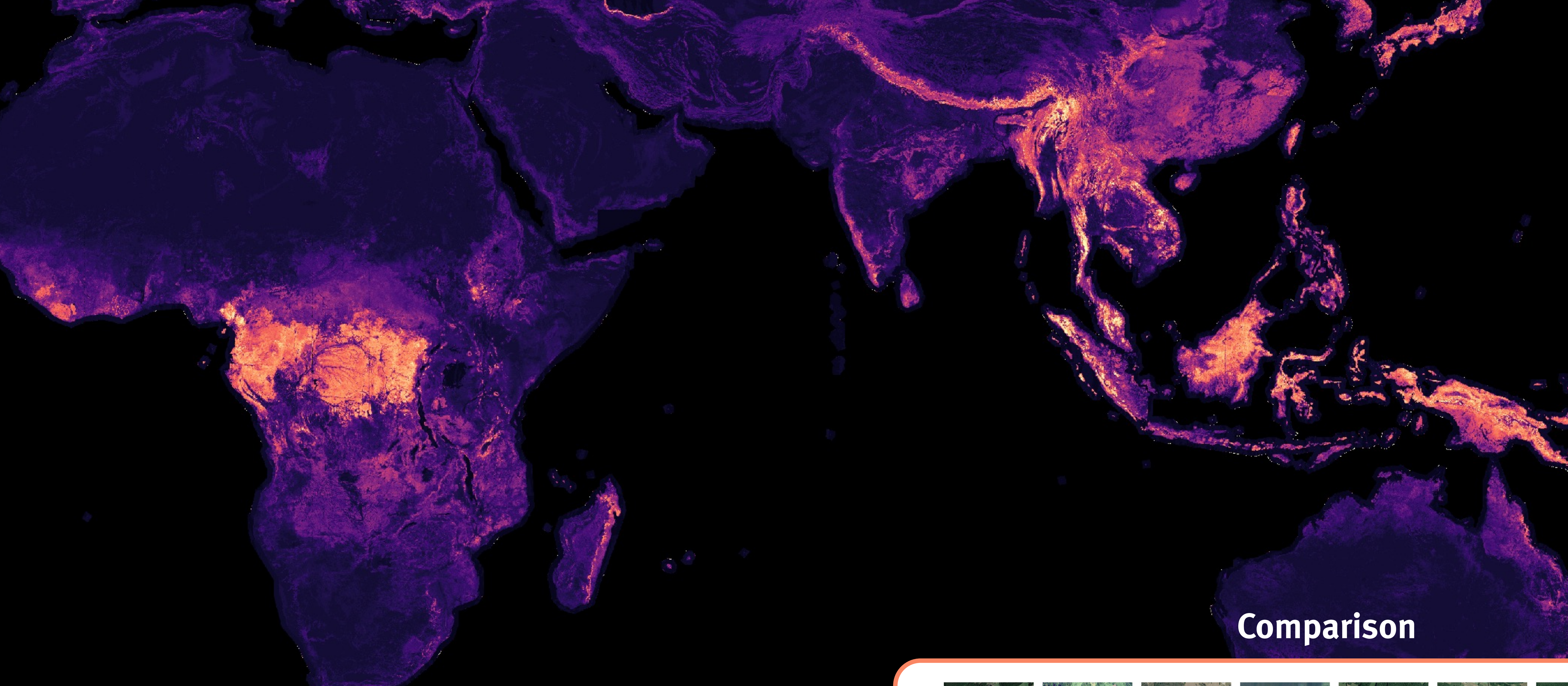
Estimating Canopy Height at Scale

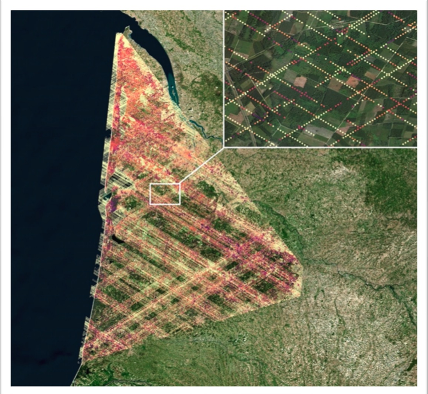
Jan Pauls[1], Max Zimmer[2], Una M. Kelly[1], Martin Schwartz[3], Sassan. Saatchi[4], Philippe Ciais[3], Sebastian Pokutta[2], Martin Brandt[5], Fabian Gieseke[1]

Universität Münster 1 · ZIB 2 · LSCE 3 · NASA 4 · 5

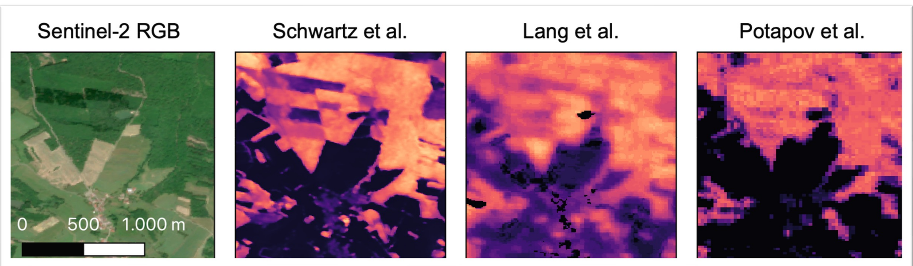


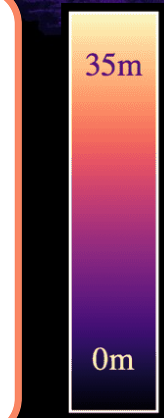
Introduction & Motivation

We propose a framework for global-scale canopy height estimation based on satellite data. Our model leverages advanced data preprocessing techniques, i.e. cloud filtering, and resorts to a novel loss function designed to counter geolocation inaccuracies inherent in the ground-truth height measurements. A comparison between predictions and ground-truth labels yields an MAE / RMSE of 2.43 / 4.73 (meters) overall and 4.45 / 6.72 (meters) for trees taller than five meters, which depicts a substantial improvement against existing global-scale maps.



A strong quality difference exists between regional (Schwartz) and global maps (Lang & Potapov). We close this gap by bringing regional map quality to a global scale.





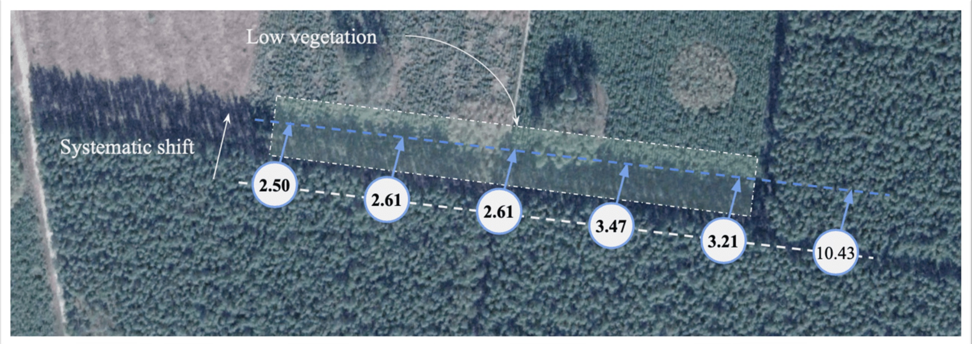
Links

Paper

GitHub

Interactive Viewer

Proposed loss



Our ground-truth GEDI height measurements are unprecise in their geolocation and clearly show a systematic shift. To combat this, we implement a shift-resilient loss function: We shift the entire track (one orbit of measurements) into all possible directions with at most a 20m magnitude. Then we select the shift with the lowest associated loss.

Define the (sparse) labels:

$$t_I(x,y) := \begin{cases} m_l & \text{if } \exists\, m \in t \text{ s.t. } (x,y)=(m_x,m_y) \\ 0 & \text{otherwise.} \end{cases}$$

$$t_I^0(x,y) := \begin{cases} 1 & \text{if } t_I(x,y) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Define the shifted labels and loss

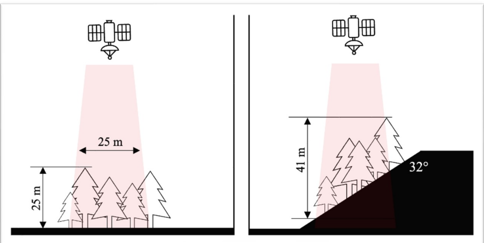
$$t_{I,\delta}(x,y) := t_I(x-\delta_x, y-\delta_y)$$

$$\mathcal{L}_S\left(f(I), T_I\right) := \frac{1}{N}\sum_{t\in T_I} \min_{\delta} \mathcal{L}\left(f(I) \odot t^0_{I,\delta}, t_{I,\delta}\right)$$

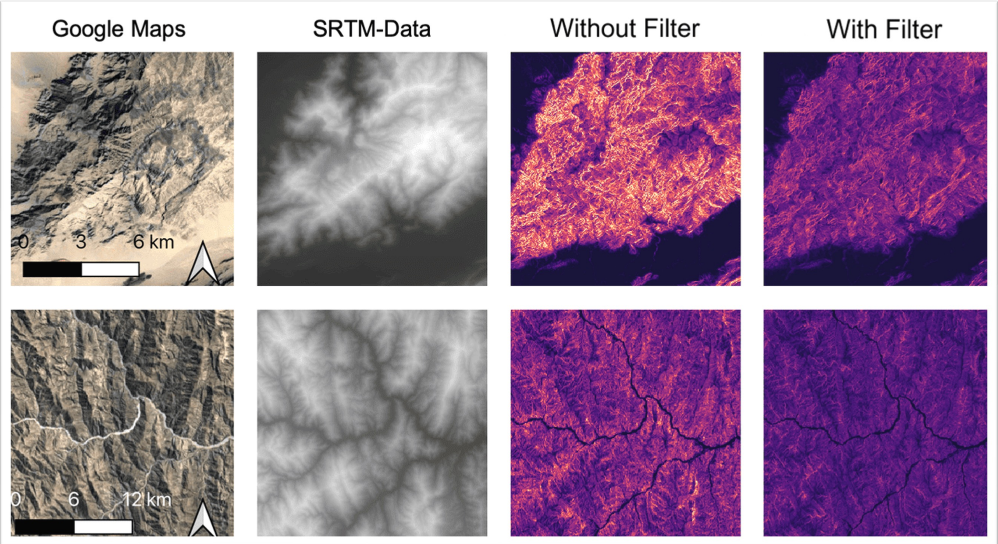
$$\text{s.t.} \quad \sqrt{\delta_x^2+\delta_y^2} \leq r.$$

Issues with Slopes

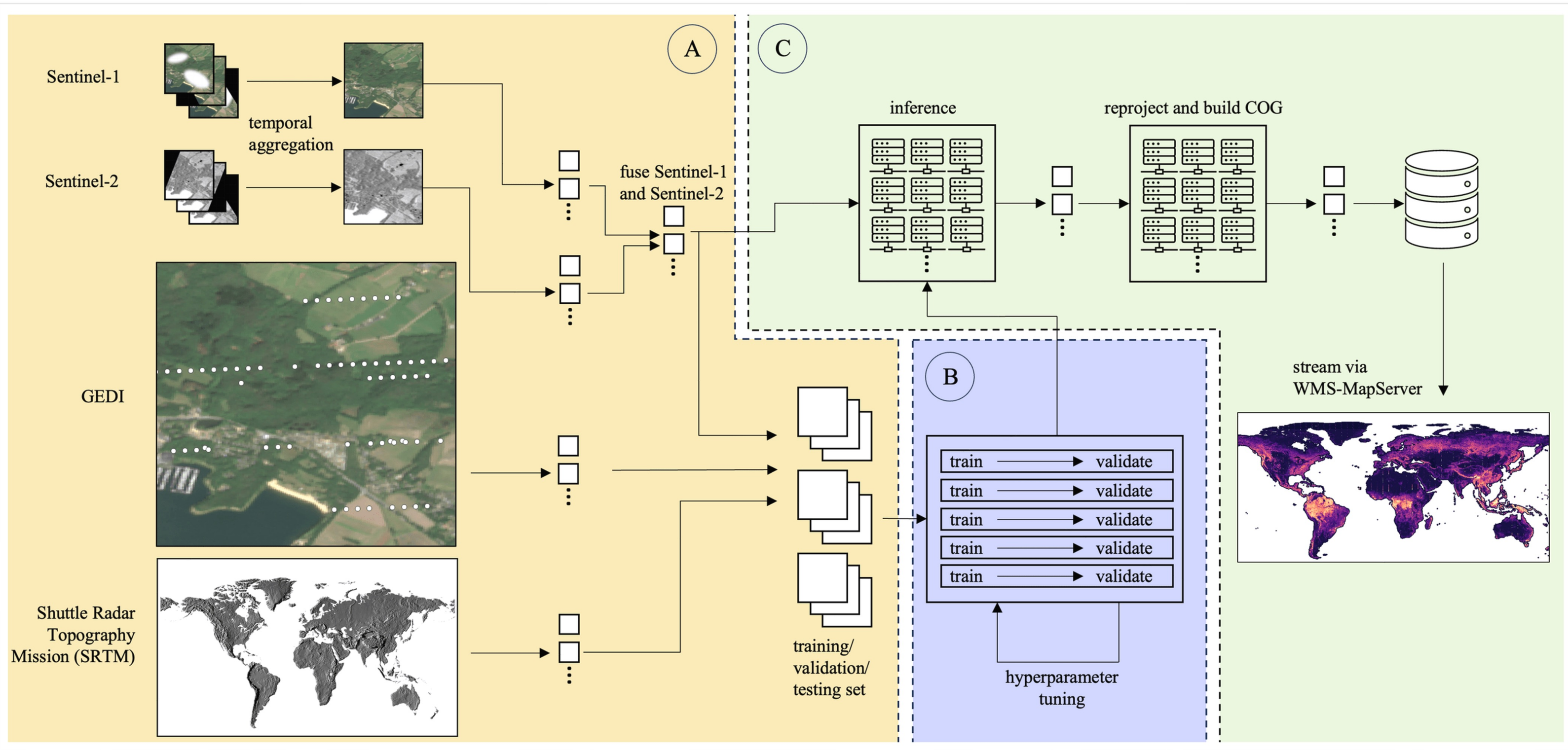
The sensor measures height by sending out photons in a 25m radius and measuring the time the photons take to return. It calculates height based on the difference between the first and last returned signals. On flat ground, this method works well. However, on sloped terrain, this approach severely overestimates the height and measured height may not accurately reflect the true height of a tree.



We use global SRTM data, to filter out ground-truth measurements on slope. As SRTM is a digital elevation model (DEM), we set a very conservative threshold (20°), to not exclude valid measurements.

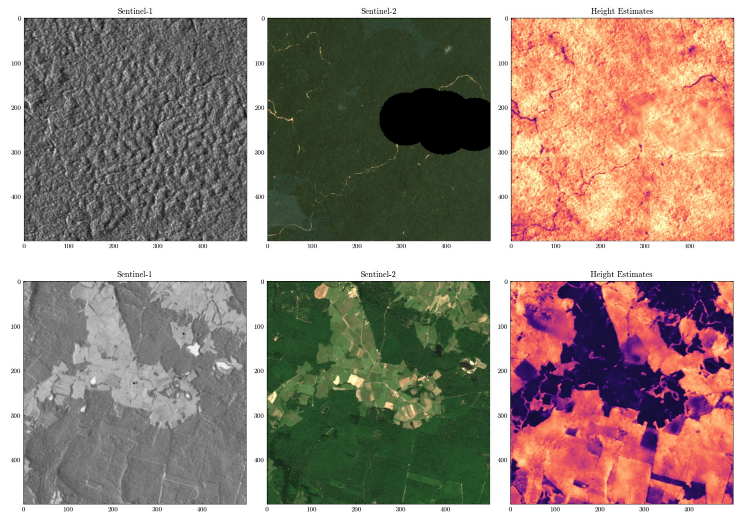


Proposed Pipeline



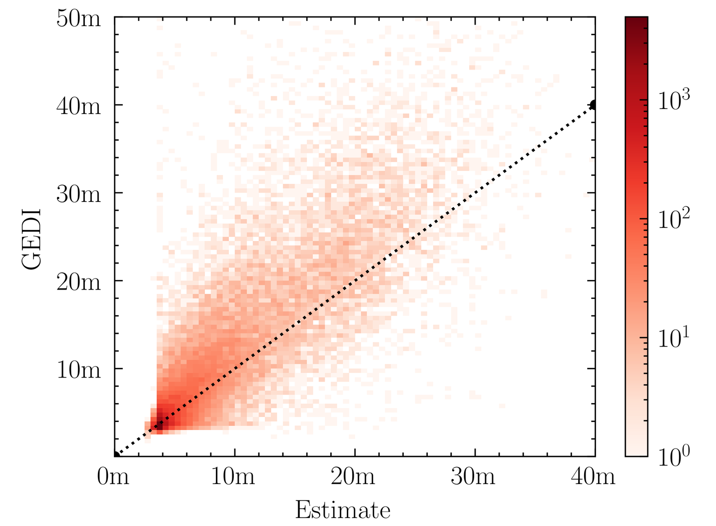
For creating a cloud-free Sentinel-1/2 image, we collect all available images from the leaf-on season, filter out clouds (only for S2), and build a per-pixel median composite. We collect all available GEDI measurements and apply several filters to increase quality, including filtering out measurements with a slope exceeding 20°. After training multiple models, we select the best one and apply it globally. To visualize the results, we build COGs (images with overviews) and stream those via a MapServer.

Results

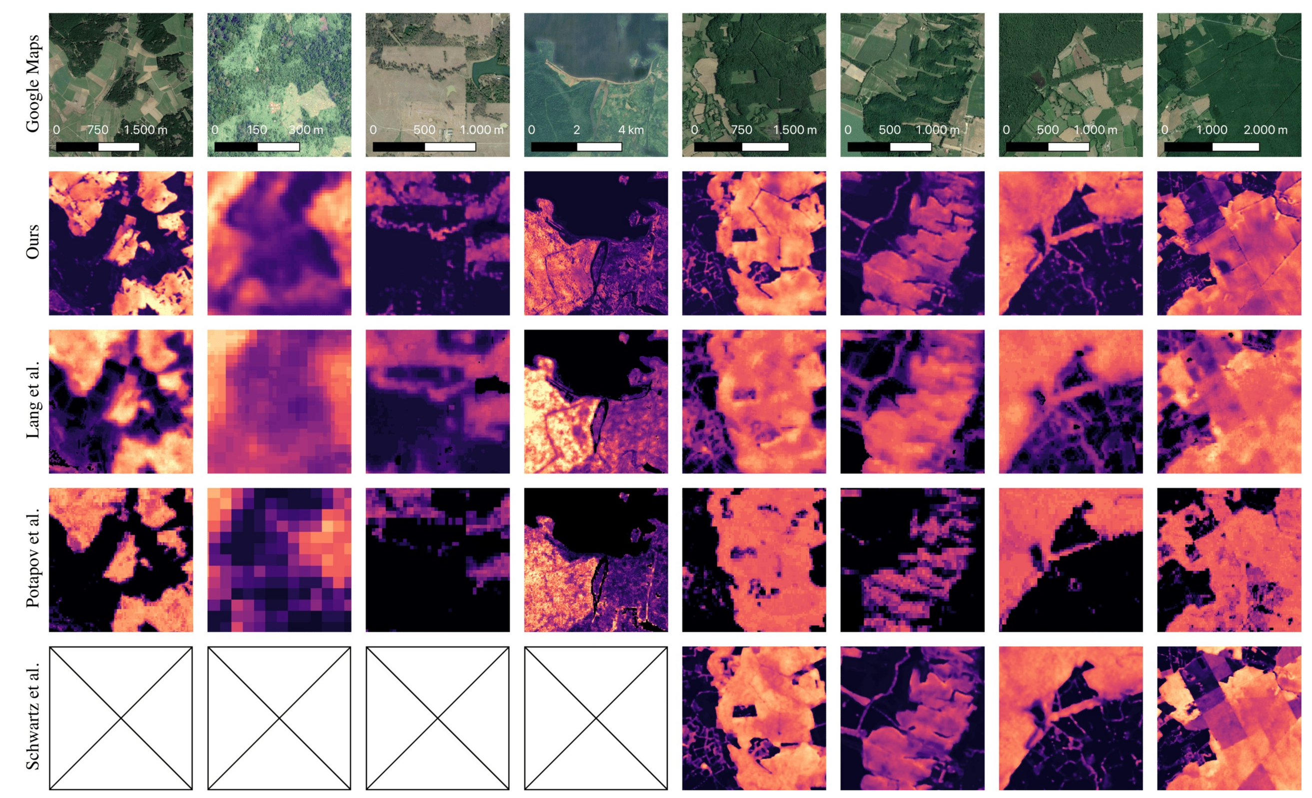


Our results show great structural details, and our model can correctly predict the height, even if the Sentinel-2 image is covered by clouds (black in the S2 image).

A scatter plot between ground-truth measurements and predictions shows a general fit, however, it also reveals a tendency to underpredict tall trees.

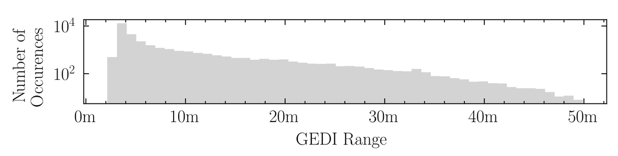


Comparison



Filter	Metric	Unit	Lang	Potapov	Ours
no	MAE	m	6.47	6.92	2.43
	MSE	m^2	74.70	85.58	22.41
	RMSE	m	8.62	9.25	4.73
	RRMSE		1.39	2.38	0.53
	MAPE		1.01	0.97	0.22
$m_l > 5$	MAE	m	8.80	10.01	4.45
	MSE	m^2	121.51	154.45	45.12
	RMSE	m	11.02	12.43	6.72
	RRMSE		1.02	1.77	0.49
	MAPE		0.76	0.77	0.28

A visual comparison between our, two other global maps and a regional one shows substantial improvements and we nearly close the gap to regional maps. When statistically comparing the maps, the improvement is consistent across multiple metrics for all trees and only trees greater than 5.



As already visible in the scatterplot, prediction performance varies greatly for different tree heights. The taller the trees get, the greater the underprediction of such trees is. Thereby, the uneven distribution of ground-truth measurements plays a key role. Still, comparing ours with two other global maps shows a clear improvement, especially regarding reduced variance.